

Repeated measures and longitudinal data

飯島勇人^{*†§}

2007 年 7 月 4 日

目次

この章で扱う内容	1
9.1 長期観測データの解析	2
9.2 繰り返し測定データの解析	3
9.3 誤差に階層性がある複数の応答変数を持つデータの解析	4

この章で扱う内容

ある単一の測定対象に関し、複数回の測定が行われているデータの解析方法。

定式化

この章で扱うような Random effect を含んだモデルは、一般に以下のように表現される。

$$y_i | \gamma_i \sim N(X_i \beta + Z_i \gamma_i, \sigma^2 \Lambda_i)$$

これは前章の式と非常によく似ているが、誤差がより一般的な共分散である Λ_i を持つ点が異なる。Random effect そのものは平均 0、分散が $(\sigma^2)D$ の正規分布で定義されるから、

$$y_i \sim N(X_i \beta, \Sigma_i)$$

という関数形となる。ここで、 $\Sigma_i = \sigma^2(\Lambda_i + Z_i D Z_i^T)$ である (Z は Random effect の入った行列、それに Λ_i を加えたもの。p. 118 や 155 を参照するとわかりやすい)。これらは各測定値に関してのものであるが、誤差と Random effect が測定値間で相関しないと仮定すると、以下のように容易に拡張できる。

$$y \sim N(X\beta, \Sigma), \Sigma = \sigma^2(\Lambda + Z\hat{D}Z^T)$$

* 北海道大学大学院農学院専門研究員

† 連絡先: hayato-i@for.agr.hokudai.ac.jp または http://www.geocities.jp/iijima_web/index.html

‡ 本ゼミのサポートページ: <http://www7.atwiki.jp/hayatoijima/pages/33.html>

§ この文章は L^AT_EX で作成しました。わーどなどではけっしてありません。

9.1 長期観測データの解析

使用データ

アメリカの個人収入の年次変動データ。個人データとして、age、educ、sex が取られている。解析の目的は、「個人年収の変動パターンに影響する要因を明らかにすること」。

データの見方

- 個人ごとに、年と収入の関係を図示する (Figure 9.1 and 9.2)
- 個人ごとの収入に対する年の回帰直線を作成し、その傾きと切片を男女間で比較する (p.186 下部の for() による算出, Figure 9.3)

これから、

- 年々収入が増加する傾向にある。
- ばらつきは大きい、女性の方が収入増加のスピードが速い

という傾向が読み取れる。これは p.188 の呪われた検定 (反応特性解析) によっても支持される。

しかし、解析方法としては、

- モデルの切片や傾きには個人差がありそうだ。
- 年変動を考慮していない。

といった、Random effect を考慮していない問題点が指摘できる。

Random effect を考慮した解析

単純のために、年が収入に与える影響が個人によって異なり、個人の影響は切片と傾き両方に現れるとするモデルを構築する。そのモデルは以下のものである。

```
> mmmod <- lmer(log(income) ~ cyear * sex + age + educ + (cyear | person), psid)
```

このモデルは、以下のように書くことができる。

$$\log(\text{income})_{ij} = \mu + \beta_y \text{year}_i + \beta_s \text{sex}_j + \beta_{ys} \text{sex}_j \times \text{year}_i + \beta_e \text{educ}_j + \beta_a \text{age}_j + \gamma_j^0 + \gamma_j^1 \text{year}_i + \epsilon_{ij}$$

i は年を示し、 j は個人 (person) を示す。

結果 (p. 189–190) の解釈としては、

- 年収は年々約 10% (8.5%) 増加している。
- age は重要な要因ではない。
- 性別の違い 1: 切片は男性が女性の 115% ほどである。
- 性別の違い 2: 交互作用項の係数が -0.0263 なので、年収の年増加率は女性が 8.5% なのに対し、男性は $0.0853 - 0.0263 = 0.059$ 、すなわち 5.9% である。

ただし、残差を見ると正規分布をしておらず (Figure 9.4) 予測値に対する残差も等分散していない (Figure 9.5) ので、income の変換方法については更なる検討が必要である。

作図蛇足

ここの部分は本文にまったく書いていなかったのですが、個人的に便利だと思ったので紹介しておきます。ある意味この章で一番役立った情報かも^^;)

カテゴリーごとの散布図として、lattice パッケージ内の `xyplot()` を使うと便利である。

```
> library(lattice)
> category <- rep(c("A", "B", "C", "D"), 25)
> tekito <- rnorm(100, 50, 10)
> darui <- tekito + rnorm(100, 10, 5)
> d <- data.frame(category, tekito, darui)
> xyplot(darui ~ tekito | category, d)
```

のように、`xyplot(y ~ x | category)`—とすると、カテゴリーごとの散布図が一気に描画される。当然だが、カテゴリーの部分は*で階層化できる。

9.2 繰り返し測定データの解析

使用データ

7 人の左右の目に光を当てたあと、目の表皮の反応のタイムラグ (100 万分の 1 秒)。なお、試験は倍率が異なるレンズごとに行われている。解析の目的は、「レンズの違いによって光を当てた後の反応までのタイムラグが異なるか？」。

データの見方

- 個人、左右の目ごとに、レンズの倍率の違いによるタイムラグを図示する (Figure 9.6)。

これらから、

- 個人ごとに傾向は異なる。
- 左右の目で傾向は異なる (しかも個人によって左右の反応に一貫性がない)

という傾向が読み取れる。そこで、Random effect としては、個人 (subject)、目の左右 (eye) があげられる。また、eye は個人間で傾向が一致しないので、nested な Random effect として考えられる。

Random effect を考慮した解析

以上の結果から、

```
> mmmod <- lmer(acuity ~ power + (1 | subject) + (1 | subject:eye), vision)
```

このモデルは、以下のように書くことができる。

$$y_{ijk} = \mu + p_j + s_i + e_{ik} + \epsilon_{ijk}$$

i は個人 (subject) j は倍率 (power) k は目の左右 (eye) を示す。

ただし、この測定は Random effect (subject かつ eye) 内で連続的に測定しているので、Random effect は測定間で相関がある可能性がある。その測定間の相関を調べる方法は、

- subject 内での相関: (subject の Std.Dev.)² / (各 Std.Dev. の 2 乗の和)
- subject の eye 内での相関: ((subject の Std.Dev.)² + (subject:eye の Std.Dev.)²) / (各 Std.Dev. の 2 乗の和)

で定義され、個人内での測定値の相関よりも、個人の目の中での測定値の相関のほうが強い (p. 193 の下部の値より)。

このモデルの有効性を、予測値-残差プロットや QQ プロットで確認している (Figure 9.7)。

9.3 誤差に階層性がある複数の応答変数を持つデータの解析

使用データ

8.8 節で用いた、ある学校のクラスごとの学力テストのデータ。8.8 節では数学のテストの結果のみを用いたが、国語 (English) の結果も同時に考慮することで、さらに有益な情報を引き出すことができる。このように、2 つ以上の応答変数画同時にあることを、Multiple response という。

Random effect を考慮した解析

*モデル式は長いので省略。結果は p. 197 に記されている。

- レファレンスカテゴリーは、「男子の国語の点数」である。
- 女子の英語の点数は、男子よりも 6.3% 高い。
- しかし数学については、交互作用 (数学の点数は女子のほうが低い) の影響により、男女の差はほとんどない。
- また、1 年次のテストの点数は今回のテストの点数と相関しているが、数学に対する影響は交互作用の影響により、英語ほどは強くない。